

Mixture model analysis of Transition Edge Sensor pulse height spectra

KEVIN J COAKLEY^{1,*}, JOLENE SPLETT¹, AND THOMAS GERRITS²

¹National Institute of Standards and Technology, Boulder CO 80305, USA

²National Institute of Standards and Technology, Gaithersburg, MD 20899, USA

*Corresponding author: kevin.coakley@nist.gov

Compiled October 25, 2021

To calibrate an optical transition edge sensor (TES), for each pulse of the light source (e.g. pulsed laser), one must determine the ratio of the expected number of photons that deposit energy and the expected number of photons created by the laser. Based on the estimated pulse height generated by each energy deposit, we form a pulse height spectrum with features corresponding to different numbers of deposited photons. We model the number of photons that deposit energy per laser pulse as a realization of a Poisson process, and the observed pulse height spectrum with a mixture model method. For each candidate feature set, we determine the expected number of photons that deposit energy per pulse and its associated uncertainty based on the mixture model weights corresponding to that candidate feature set. From training data, we select the optimal feature set according to an uncertainty minimization criterion. We then determine the expected number of photons that deposit energy per pulse and its associated uncertainty for test data that is independent of the training data. Our uncertainty budget accounts for random measurement errors, systematic effects due to mismodeling feature shapes in our mixture model, and possible imperfections in our feature set selection method. © 2021 Optical Society of America

<http://dx.doi.org/10.1364/ao.XX.XXXXXX>

1. INTRODUCTION

In our experiment, both background photons and photons produced by a pulsed-laser source deposit energy in a Transition Edge Sensor (TES) [1, 2]. Each energy deposit generates a waveform from which we determine a pulse height. We attribute features in the pulse height spectrum to a blackbody radiation background process as well as energy deposits of photons created by the pulsed-laser source. We assume that other background sources such as cosmic rays are negligible. Our primary goal is to estimate the expected number of photons that deposit energy per pulse. This expected value is less than the expected value of photons created by each pulse of the laser because the probability that each photon creates an energy deposit is less than 1. To calibrate a TES, one must determine its photon detection efficiency which is the ratio of these expected values. One can estimate the expected value of the number of photons created by each laser pulse and its associated uncertainty with methods from [3]. In this work, we present new methods to estimate the expected number of photons deposited in the TES per pulse, and its associated uncertainty based on mixture model analysis of TES pulse height spectra. Given these mixture model analysis results, and earlier analyses based on methods from [3], one can determine the photon detection efficiency of a TES and its associated uncertainty. Here, we focus solely on the mixture

model analysis that can be implemented in the TES calibration procedure.

Many researchers have measured TES detection efficiency and performed detector tomography [4–7]. In [4], the mean number of photons deposited in the TES and its associated random uncertainty (but not systematic uncertainties) were determined by fitting multiple Gaussian peaks to pulse height spectra where the amplitude of each peak was a product of a constant and a Poisson probability term. Here, we determine the mean number of photons deposited by a pulsed laser with a mixture model approach and quantify components of uncertainty due to both random effects and systematic effects. We also select feature sets based on estimated mixture model weights determined from the feature according to an uncertainty minimization criterion. In Section 3.D, we demonstrate that our feature selection fails if we replaced our mixture model weights with weights determined from a multiple Gaussian fit like in [4].

In Section 2, we discuss experimental details. In Section 3 we discuss our method for determining a pulse height from a waveform produced by energy deposits in the TES, a mixture model [8, 9] method to determine the expected number of pulsed-laser photons per pulse that deposit energy in the TES, and a method to select the optimal set of features for determining the expected number of photons that deposit energy in the TES. In Section

4, we fit mixture models to pulse height data acquired from two different experiments performed at the National Institute of Standards and Technology (NIST) and estimate the expected number of photons per pulse that deposit energy in the TES. The quality of the data from the two different experiments varies dramatically. For each experiment, we quantify components of uncertainty due to random measurement errors, imperfect modeling of feature shapes, and possible imperfect performance of our feature set selection method which has the potential to reduce uncertainty in general applications. For discussion of different but related mixture model analyses of TES spectra, see [10, 11].

2. EXPERIMENTAL DETAILS

Optical TESs are energy-resolving single-photon detectors. When a TES is illuminated with a laser pulse with a random number of photons of the same energy, one can determine the expected number of photons in the pulse [4, 12]. In general, for each pulse, the observed pulse height is a realization of a random variable with a distribution that depends on the unobserved number of photons deposited by the laser. Further, spectral features corresponding to different numbers of deposited photons overlap. [10, 13, 14]. Typically, a sum of Gaussians are fit to the pulse height spectrum where the amplitude of each Gaussian is a product of a constant and a Poisson probability term. Given a sufficiently accurate estimate of the incident photon rate, one can determine detection efficiencies as high as 0.98 [4, 7], and their associated uncertainties and/or coverage intervals. For the case where multiple-photon energy deposits occur with negligible probability, the expected number of photons that deposit energy in a TES can be determined in 'click' detector experiments where one detects an electrical pulse with a shape that is independent of the number of absorbed photons [15].

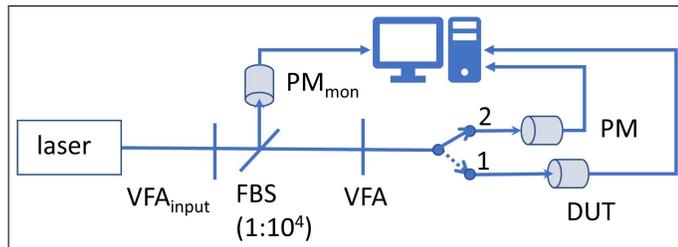


Fig. 1. Schematic of measurement system. The combination of a fiber beam splitter (FBS) and variable fiber attenuator (VFA) establishes a reliable method to predict the output power at the device under test (DUT) by measuring the power ratio at both fiber outputs with two optical fiber power meters (PM and PM_{mon}). The input VFA (VFA_{input}) is used to attenuate and adjust the photon rate at the DUT.

Our experiment setup (see Figure 1) is the same as the experimental setup described in [3]. In our experiment, a fiber-coupled, pulsed-diode laser produces photons in a weak coherent state with wavelength (energy) of 1553 nm (0.8 eV) at a repetition rate of 50 kHz and a pulse width and mean power of approximately 100 ns and 2 μ W respectively. The laser pulses are directed to a variable fiber attenuator (VFA_{input}) and a fiber beam splitter (FBS). The high-power output port is directed to a fiber-coupled monitor power meter (PM_{mon}). The low-power output is directed to a second variable fiber attenuator (VFA). With the VFA,

one can vary the ratio of these powers. Determination of the photon rate at the device under test (DUT), in our case a TES, is done as follows. With the VFA_{input} at the lowest attenuation setting, the individual power on PM_{mon} and PM are recorded. These power measurements determine the monitor-to-output ratio between PM_{mon} and the end of the fiber at PM. To measure the photon rate at the DUT, we increase the attenuation of VFA_{input} so that photon rates on the order of 1 photon per laser pulse at the end of the fiber through monitoring the laser power and applying the measured monitor-to-output ratio. We then splice the DUT fiber to the fiber exiting from the FBS. The rate at which our data acquisition system records the output waveforms produced by the TES is 10 million samples per second. The TESs, with a critical temperature of approximately 170 mK, were optimized for an operating wavelength of 1550 nm and operated in an adiabatic demagnetization refrigerator at a constant temperature of 100 mK. We present data for two different TESs. In Experiment A, the TES acquired high signal-to-noise data. In Experiment B, the TES acquired low signal-to-noise data.

3. ANALYSIS METHODS

A. Pulse Height Estimation

In our pulsed-laser experiment, we measure a waveform and the associated pulse height generated by each pulse. We determine each pulse height as the difference between the maximum value of the waveform and the mean value of the five earliest values of the waveform. We report results for two experiments performed under different conditions – Experiment A and Experiment B. For Experiment A, we report results for two cases. In the primary analysis, outlier waveforms are excluded from the analysis. In a second analysis, all waveforms are analyzed. We denote the data analyzed for these two cases as the “filtered” data and the “unfiltered” data. We attribute outlier waveforms to random photon energy deposits not produced by our pulsed-laser source. We also report results for unfiltered Experiment B data.

B. Mixture Model

We model the pulse height probability density function (pdf) with a mixture model [8, 9]. The mixture model pdf, f_{mix} , is

$$f_{mix}(x) = \sum_{k=0}^{k_{max}} w_k f(\theta_k, x), \quad (1)$$

where $f(\theta_k, x)$ is the pdf for an event due to the deposit of k photons from the pulsed-laser source. The weights (w_k where $k=0,1,2,3, \dots, k_{max}$) are non-negative and sum to 1. That is,

$$\sum_{k=0}^{k_{max}} w_k = 1. \quad (2)$$

We determine the parameters of the mixture model and their weights based on unbinned pulse height data, (x_1, x_2, \dots, x_N) where N is the total number pulse heights, with a maximum likelihood method. In this approach, we maximize the log-likelihood function of the data, $\log L$, where

$$\log L = \sum_{i=1}^N \log (f_{mix}(x_i)), \quad (3)$$

as a function of the weights and the distributional parameters (under the assumption that measured pulse heights are realizations of independent random variables).

In our studies, k_{max} is no larger than 5. For each case of interest, we represent the observed pulse height data as a histogram. For each bin of the histogram, we predict the expected number of counts as the product of the integral (from the lower to the upper endpoint of the bin) of $f_{mix}(x)$ and a scale factor. We choose the scale factor so that the sum (over all bins) of the observed counts and the sum of the predicted counts in the histogram agree.

In this work, we consider both a normal (Gaussian) mixture model and a gamma mixture model. In the Gaussian mixture model, $\theta_{bg} = (\mu_b, \sigma_b)$ where μ_b and σ_b are the mean and standard deviation of Gaussian random variable. For a k -photon energy deposit, $\theta_k = (\mu_k, \sigma_k)$. The pdf for a Gaussian random variable is, for $-\infty \leq x \leq \infty$,

$$f(\mu, \sigma, x) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right). \quad (4)$$

For a gamma random variable, the associated pdf is, $0 \leq x \leq \infty$,

$$f(\beta, \gamma, x) = \frac{1}{\beta\Gamma(\gamma)} \left(\frac{x}{\beta}\right)^{\gamma-1} \exp\left(-\frac{x}{\beta}\right), \quad (5)$$

where $\gamma > 0$ and $\beta > 0$, and the Γ function is defined as

$$\Gamma(a) = \int_{t=0}^{\infty} t^{a-1} \exp(-t) dt. \quad (6)$$

For each model, we determine the weights and the distributional model parameters by maximizing the likelihood function of the unbinned measured pulse heights with the expectation-maximization (EM) algorithm [16]. In particular, we implement the EM algorithm for the normal and gamma models with the public domain R [17] functions “normalxEM” and “gam-mamixEM” from the mixtools package [18]. We halt the EM algorithm when the change of the log-likelihood at successive iterations is less than 10^{-8} . Initial values for the mean values of the pdfs in the mixture are determined from the observed features with a peak-finding function [19]. Initial values for other parameters that determine the standard deviation of the pdfs are based on the analysis of the dispersion of each feature in the pulse height spectra.

C. Estimation of expected number of photons deposited per pulse

We model the number of photons created per pulse as a Poisson random variable with expected value θ_c . Not all the created photons deposit energy in the TES. Photons can be lost in transit from the laser to the TES. Also, it is possible for a photon to scatter off the TES detector without depositing energy. Based on physical considerations, whether a particular photon deposits energy or not in the TES has no influence on whether any other photon does or does not. Because of this independence, if N_c photons are created by the laser, and p_{dep} is the probability that a photon deposits its energy in the TES, we model the number of photons that deposit energy as a binomial random variable with expected value $N_c p_{dep}$ where $0 \leq p_{dep} \leq 1$. It follows that the number of photons that deposit energy per pulse is a realization of a Poisson process with expected value $\theta_{dep} = p_{dep}\theta_c$.

Our estimate of θ_{dep} is based on the estimate of the weights associated with the features generated by photon energy deposit events. Given the theoretical weight for a j -photon deposit feature, $w(j)$, we define a normalized weight,

$$w_*(j) = \frac{w(j)}{\sum_j w(j)}, \quad (7)$$

where the j index denotes features that are included in the analysis to determine θ_{dep} . For instance, if we include all the photon energy deposit weights, then j would range from 0 to k_{max} . However if we exclude the last weight from the analysis, then j would range from 0 to $k_{max} - 1$.

Here, we assume that the number of photons in a pulse is a realization of Poisson random variation with expected value θ_{dep} . Based on this assumption, we have

$$w_*(j) = f(j, \theta_{dep}) = \frac{\Pr(j|\theta_{dep})}{\sum_j \Pr(j|\theta_{dep})}, \quad (8)$$

where the probability that j photons deposit energy is

$$\Pr(j|\theta_{dep}) = \frac{\exp(-\theta_{dep})\theta_{dep}^j}{j!}. \quad (9)$$

Given that the mixture model analysis estimate of the theoretical value $w_*(j)$ is $\hat{w}_*(j)$, we estimate θ_{dep} by the method of nonlinear least squares by minimizing

$$\sum_j (\hat{w}_*(j) - f(j, \theta_{dep}))^2 \quad (10)$$

as a function of θ_{dep} .

D. Feature selection

In our analysis of experimental results, we are free to determine θ_{dep} from weights determined from all the features due to photon energy deposits or any subset of these features. We select the optimal subset for analysis according to an uncertainty minimization criterion. To guard against overly optimistic uncertainties for θ_{dep} , we identify the optimal subset from training data that does not overlap with the primary data of interest. We denote this primary data as the test data. We randomly split the observed pulse height data into training data and test data with a resampling without replacement method. The N measured pulse heights are indexed according to their observation time. The resampling without replacement method randomly permutes the order of the set of integers $(1, 2, 3, \dots, N)$. We simulate a permutation of the set of integers with the function “sample” in R [17]. Pulse heights with an associated index in the first half of the permuted integers are assigned to the training data. The remaining pulse heights are assigned to the test data. To illustrate our method, consider the simple case of four measurements x_1, x_2, x_3, x_4 acquired at times t_1, t_2, t_3, t_4 where $t_1 < t_2 < t_3 < t_4$. If the random permutation of $(1, 2, 3, 4)$ were $(2, 4, 3, 1)$, the training data would be (x_2, x_4) , and the test data would be (x_3, x_1) .

We determine θ_{dep} and its associated uncertainty from the test data based on the feature subset identified from the training data. One could determine weights based on the fitting approach in [4]. However, if we replaced our mixture model weights with weights determined from fitting methods like [4] in Eq. 10, all feature subsets would yield the same estimate of θ_{dep} and the same estimate of its associated random uncertainty. Hence, our feature selection method is not feasible if based on weights determined from the approach in [4]. This follows from the observation that if weights are determined from the [4] fitting approach, the estimated normalized weight (see Eq. 7) for the k th feature is $A^{-1}H(k) \frac{\exp(-\hat{\theta}_{dep})(\hat{\theta}_{dep})^k}{k!}$ where $\hat{\theta}_{dep}$ is our estimate determined from the fit to full pulse height spectrum, and

$$A = \sum_{k=0}^{k_{max}} H(k) \frac{\exp(-\hat{\theta}_{dep})(\hat{\theta}_{dep})^k}{k!}, \quad (11)$$

where $H(k) = 1$ if k is in the feature set, and $H(k) = 0$ if k is not in the feature set.

4. RESULTS

A. Experiment A: high data quality

For Experiment A, we analyze pulse heights determined from filtered data where outlier waveforms are excluded from the analysis. We identify a waveform as an outlier if the first time for which the waveform exceeds -25 mV is less than $0.5 \mu\text{s}$ or larger than $1.15 \mu\text{s}$ after the expected arrival time of the pulse. The threshold of -25 mV corresponds to an estimate of the mean baseline (-62.3 mV) plus 3.2 times the standard uncertainty of the measured baseline. The fraction of waveforms that are identified as outliers is 0.022. We illustrate our acceptance/rejection method for a small number (400) of the total number (1,630,104) of waveforms acquired in Figure 2.

In Table 1, we show estimated model parameters determined from the training data for a normal mixture model. In addition to a "0-photon" feature to due to background photons and no-photon events, there are five other features corresponding to energy deposits of k -photons (deposited by the pulsed laser) where $k = 1, 2, \dots, 5$. We excluded a seventh feature from the analysis because it had a very low signal-to-noise ratio. For each feature, we show the estimate of the mean μ and standard deviation σ of the associated normal distribution, and the corresponding weight for the feature. The standard errors of the parameter estimates are obtained by nonparametric bootstrap resampling [20] of the observed pulse heights.

Table 1. Normal mixture model results for filtered Experiment A data. Bootstrap standard errors (components of uncertainty due to random measurement errors) are shown in parentheses for each estimate. For instance, 0.22006(45) means that the estimate and its associated bootstrap standard error are 0.22006 and 0.00045 respectively.

Feature	μ (mV)	σ (mV)	Weight
Training Data			
0-photon	19.90(03)	10.87(02)	0.22006(45)
1-photon	89.79(03)	15.02(03)	0.33812(53)
2-photon	165.71(04)	14.89(03)	0.25308(51)
3-photon	233.14(06)	14.50(05)	0.12652(41)
4-photon	292.32(10)	14.24(11)	0.04899(29)
5-photon	343.62(17)	11.28(10)	0.01323(15)
Test Data			
0-photon	19.88(03)	10.91(02)	0.22080(49)
1-photon	89.73(03)	15.03(03)	0.33802(53)
2-photon	165.74(04)	14.95(03)	0.25204(49)
3-photon	233.26(06)	14.46(05)	0.12720(40)
4-photon	292.71(10)	14.45(12)	0.04885(29)
5-photon	343.92(18)	11.12(11)	0.01308(15)

In Table 2, we list estimates of θ_{dep} determined for different choices of the features analyzed for the filtered training data. For

Table 2. Six-feature, normal mixture model results for filtered Experiment A data. Here, we estimate θ_{dep} for different subsets of the features. The nomenclature for the feature set denotes the range of features from which we determine θ_{dep} . For instance, the feature set "2-4" corresponds to results based on analysis of the weights associated with 2-photon, 3-photon and 4-photon features. We determine θ_{dep} from the test data for the feature set that yields the lowest value of u_{subtot} for the training data. From the test data, we also estimate θ_{dep} from the feature set that yields the second lowest value of the combined uncertainty u_{subtot} for the training data. From the two estimates of θ_{dep} for the test data, we determine a component of uncertainty due to imperfect performance of our feature selection method. Uncertainty due to imperfect feature selection, $u_{feature}$, for the training and test data are 0.0017 and 0.0022 respectively.

Feature set analyzed	θ_{dep} Normal Model	u_{ran}	θ_{dep} Gamma Model	u_{model}	u_{subtot}
Training Data					
2-5	1.5048	0.0039	1.5047	0.0001	0.0039
2-4	1.5108	0.0044	1.5083	0.0014	0.0046
3-5	1.5132	0.0076	1.5090	0.0024	0.0080
1-5	1.5001	0.0022	1.5146	0.0083	0.0086
1-4	1.5018	0.0023	1.5163	0.0083	0.0086
0-5	1.5101	0.0015	1.4953	0.0086	0.0087
0-4	1.5114	0.0016	1.4953	0.0093	0.0094
0-3	1.5107	0.0018	1.4924	0.0106	0.0107
1-3	1.4978	0.0028	1.5159	0.0104	0.0108
0-2	1.5153	0.0024	1.4859	0.0170	0.0172
Test Data					
2-5	1.5114	0.0037	1.5141	0.0016	0.0040
2-4	1.5191	0.0042	1.5182	0.0005	0.0042

each set of features chosen for analysis, we determine estimates of θ_{dep} by nonlinear least squares as described in Section 3.C for both the normal and the gamma mixture models. For the normal mixture model, we determine the component of uncertainty of θ_{dep} due to random effects, u_{ran} , with the nonparametric bootstrap method. In a separate analysis, we determine a component of uncertainty due to imperfect knowledge of the shape of the features in the mixture model, u_{model} , based on the difference between estimates of θ_{dep} determined by the normal mixture model and a gamma mixture model. Following [21], we quantify this component as the standard deviation of uniform distribution with width equal to twice the magnitude of the difference of the estimates. Thus, if the magnitude of the difference of the two estimates is δ , $u_{model} = 2\delta/\sqrt{12}$. We sort results (see Table 2) determined from the training data according to the values of a combined uncertainty u_{subtot} where

$$u_{subtot} = \sqrt{u_{ran}^2 + u_{model}^2}. \quad (12)$$

We identify the optimal feature set as the one that yields the lowest value of u_{subtot} determined from the training data. We determine θ_{dep} from the test data based on this optimal feature set.

We determine a third component of uncertainty, $u_{feature}$, from the test data that accounts for possible imperfection of our feature set selection method. We first determine θ_{dep} from the test data for the two feature sets that yield the lowest and second lowest values of u_{subtot} from the training data. We determine $u_{feature}$ as twice the standard deviation of a uniform distribution with lower and upper bounds equal to the smaller and large values of these two estimates. Given $u_{feature}$, we compute a combined standard uncertainty u_{tot} as

$$u_{tot} = \sqrt{u_{subtot}^2 + u_{feature}^2}. \quad (13)$$

From the filtered Experiment A training data, the selected feature sets that yield the two lowest values of u_{subtot} are 2-5 and 2-4 (see Table 2.) Based on these two feature sets, for the test data, our estimate of θ_{dep} is 1.5114, $u_{subtot} = 0.0040$, $u_{feature} = 0.0022$, and hence $u_{tot} = 0.0046$ (see Table 2). We note that if we had switched the roles of the training and test data in our study, the resulting sorted order of the feature sets that yield the two lowest values of u_{subtot} determined from the training data would be the same.

In Figures 3 and 4, for the “filtered” test data set, we compare the observed pulse height spectrum to predicted contributions from each of the six features determined by a normal mixture model analysis and a gamma mixture model analysis. In Figure 5, we compare the observed pulse height spectrum to the overall prediction (sum of the predictions from all six features).

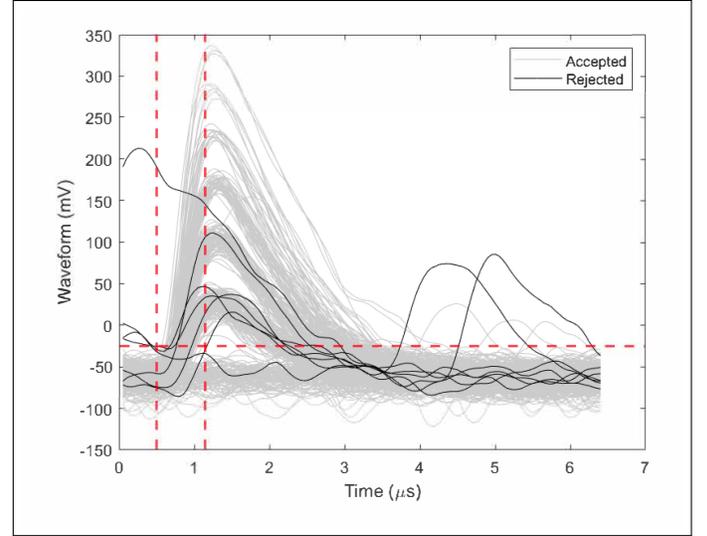


Fig. 2. For the first 400 waveforms acquired in Experiment A, we show accepted and rejected (outlier) waveforms. Discrimination thresholds (see Section 4.A) that define acceptance and rejection regions are shown as dashed red lines. The times shown on x-axis are relative to a reference trigger time.

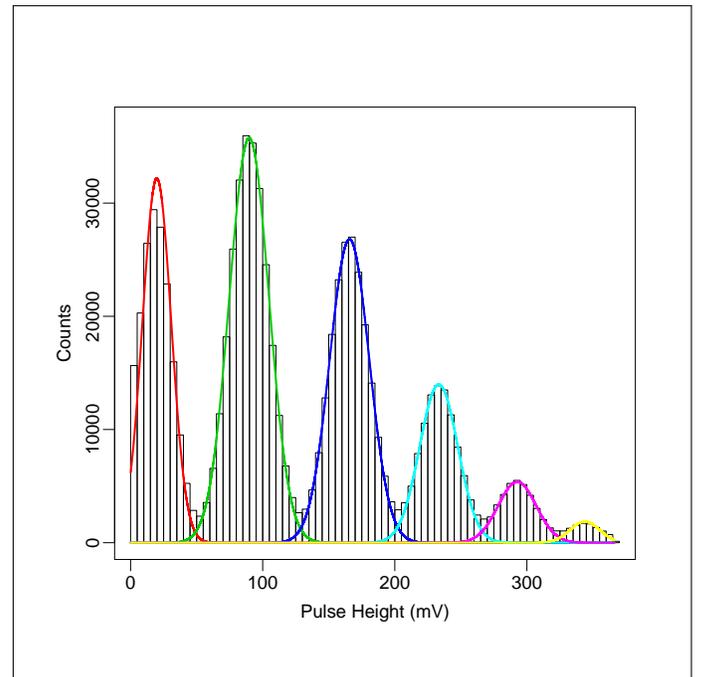


Fig. 3. For filtered test data from Experiment A, we compare the observed pulse height spectrum with the predicted contributions corresponding to each of six features determined for a normal mixture model analysis. The overall prediction (not shown) is the sum of the six predictions shown here.

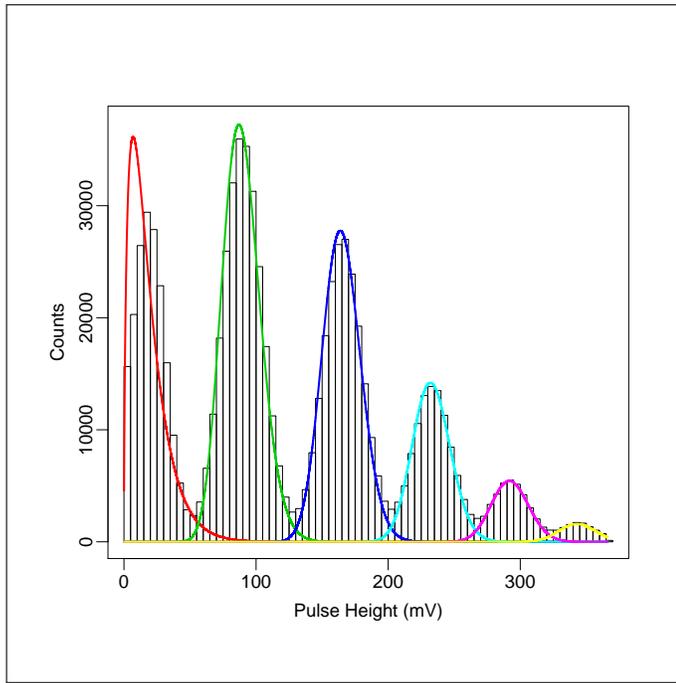


Fig. 4. For filtered test data from Experiment A, we compare observed pulse height spectrum with the predicted contributions corresponding to each of six features determined by a Gamma mixture model analysis. The overall prediction (not shown) is the sum of the six predictions shown here.

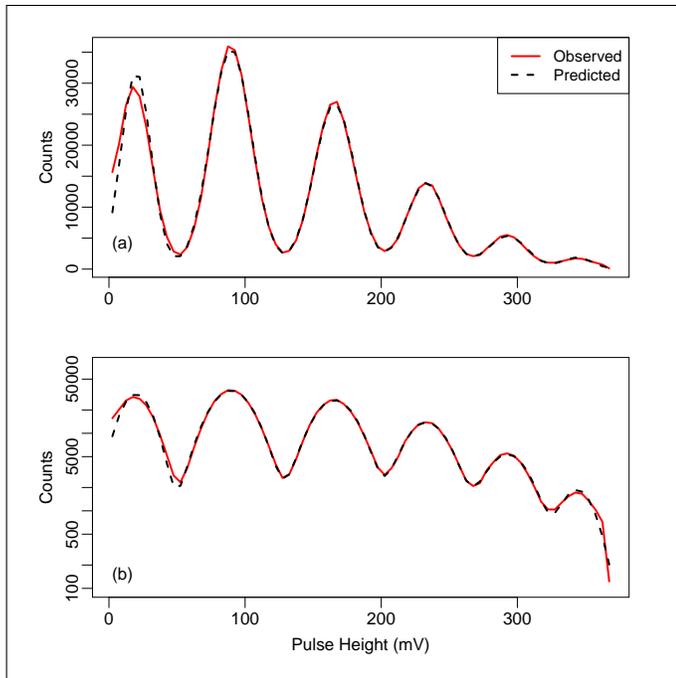


Fig. 5. For filtered test data from Experiment A, we compare observed and overall predicted pulse height spectra. The predicted spectrum is determined with a normal mixture model. In (a) and (b), we plot count data on a linear scale and a log scale respectively.

B. Experiment B: low data quality

We analyze “unfiltered” data from Experiment B. That is, all observed waveforms are analyzed. We excluded a sixth feature from the analysis because it had a very low signal-to-noise ratio. In all, we acquired 2,092,132 waveforms. In the inset of Figure 6b, we display 400 representative waveforms. We did not develop a method and associated algorithm to identify and reject possible outlier waveforms in Experiment B. Whether Experiment B produced outlier waveforms, and whether any such outliers significantly affected results may be worthy of further study.

For the normal mixture model, we show estimates of μ and σ and the corresponding weight for each feature in Table 3 for the training data. In Table 4, we show estimates of θ_{dep} for feature sets sorted according to u_{subtot} (from lowest to highest). We select “2-4” as the optimal feature set because it yields the lowest value of u_{subtot} for the training data.

Table 3. Five-feature normal mixture model results for unfiltered pulse height data corresponding to Experiment B data. Bootstrap standard errors are shown in parentheses.

Feature	μ (mV)	σ (mV)	Weight
Training Data			
0-photon	4.629(11)	2.337(06)	0.29460(94)
1-photon	13.945(11)	3.353(12)	0.44021(112)
2-photon	25.776(11)	2.622(11)	0.17815(66)
3-photon	35.399(23)	3.215(39)	0.07259(69)
4-photon	44.364(66)	2.528(29)	0.01444(31)
Test Data			
0-photon	4.611(11)	2.332(05)	0.29387(92)
1-photon	13.927(11)	3.379(12)	0.44125(112)
2-photon	25.767(12)	2.607(11)	0.17667(70)
3-photon	35.344(25)	3.288(43)	0.07408(74)
4-photon	44.435(73)	2.495(31)	0.01413(33)

Given the selected feature set “2-4” determined from the training data, for the test data we determine θ_{dep} to be 1.1682 and $u_{model} = 0.0150$. Based on the two feature sets that yield the two lowest values of u_{subtot} for the training data, we obtain $u_{feature} = 0.0850$ and $u_{tot} = 0.0868$ for the test data (see Table 4.) Like in Experiment A, if we had switched the roles of the training and test data in our study, the feature sets that yield the two lowest values of u_{subtot} from the training data would be the same. In Figure 6 we compared the observed pulse height spectrum to the overall normal mixture model prediction due to all five features.

5. DISCUSSION

A. Optimal feature set

For both Experiment A and Experiment B, neither the 0-photon nor the 1-photon feature appear in the selected feature set (see Tables 2 and 4). Visual evidence (see Figures 3, 4, 5, and 6) suggests that the 0-photon feature is not as well-predicted (compared to other features) by either the gamma model or the

Table 4. Five-feature, normal mixture model results for unfiltered Experiment B data. Here, we estimate θ_{dep} for different subsets of the photon features. For the test data $u_{feature}$ is 0.0850.

Feature set analyzed	θ_{dep} Normal Model	u_{ran}	θ_{dep} Gamma Model	u_{model}	u_{subtot}
Training Data					
2-4	1.1473	0.0083	1.1327	0.0084	0.0118
1-3	0.8759	0.0031	0.9078	0.0184	0.0187
1-4	0.8814	0.0028	0.9161	0.0200	0.0202
0-4	1.1369	0.0015	1.0852	0.0299	0.0299
0-3	1.1465	0.0017	1.0895	0.0329	0.0330
2-3	1.2225	0.0150	1.1686	0.0311	0.0345
0-2	1.1700	0.0022	1.1018	0.0394	0.0395
3-4	0.7958	0.0233	0.9567	0.0929	0.0958
Test Data					
2-4	1.1682	0.0088	1.1423	0.0150	0.0174
1-3	0.8737	0.0031	0.9102	0.0211	0.0213

normal model. Perhaps this is why our feature selection method excludes the 0-photon feature. Maybe the 1-photon feature is not selected because difficulty in modeling the 0-photon feature affects the accuracy of the mixture model prediction of the 1-photon due to overlap effects. The order of feature sets that yield the two lowest values of u_{subtot} was the same for both the training and test data. This suggests that our feature set selection method is sound.

B. Experiment A: comparison of results filtered and unfiltered data

For the filtered Experiment A data, we determine θ_{dep} to be 1.5114(46) (see Table 2 and Section 4.A) In an additional study, we determine results for unfiltered Experimental A data. For both the filtered data and unfiltered data, the feature set that yields the minimum value of $u_{subplot}$ for the training data, and the second lowest value of $u_{subplot}$ for the training data, are "2-5" and "2-4" respectively. Thus, the optimal feature set determined by our methods is the same for the both unfiltered and filtered Experiment A data. From the test data, for the unfiltered data, we determine θ_{dep} to be 1.5080(45). For completeness, for the unfiltered Experiment A data, $u_{ran} = 0.0037$, $u_{model} = 0.0012$, $u_{subtot} = 0.0039$ and $u_{feature} = 0.0022$. The fraction of waveforms identified as outliers and excluded from the filtered data is 0.022. For general applications, as the fraction of outliers increases, we expect that the magnitude of the difference between the estimates of θ_{dep} determined from filtered data and unfiltered data to increase.

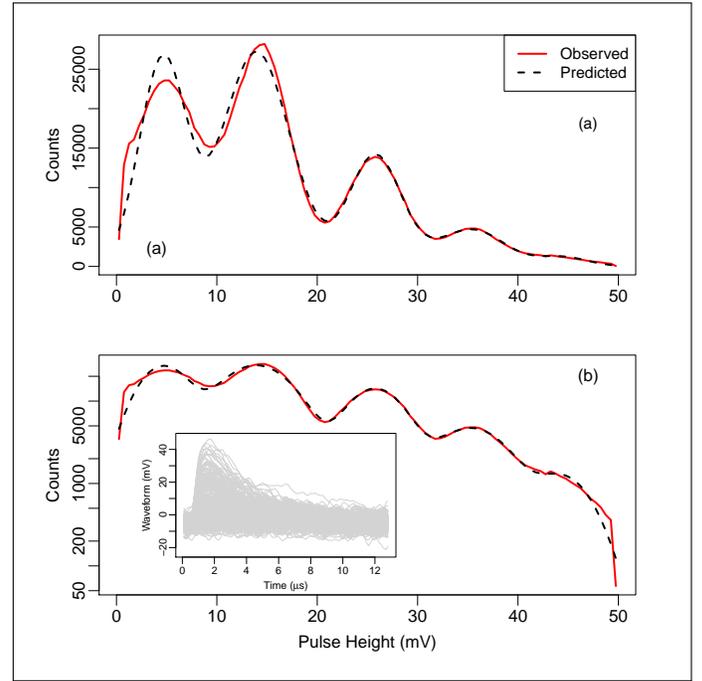


Fig. 6. We compare the observed pulse height spectrum and the overall predicted pulse height spectrum determined with a normal mixture model for Experiment B test data. In (a) and (b), we plot count data on a linear scale and on a log scale respectively. The inset plot in Figure 6b displays 400 representative waveforms.

C. Experiment B

For Experiment B, the variation of the estimates of θ_{dep} for different sets of features chosen for analysis is greater than what we observed for Experiment A. Suppose that we define the fractional range, of estimates of θ_{dep} over all subsets as

$$\left(\max(est) - \min(est) \right) / \left(\left(\max(est) + \min(est) \right) / 2 \right),$$

where $\max(est)$ and $\min(est)$ are the maximum estimate of θ_{dep} and the minimum estimate of θ_{dep} respectively. For the unfiltered Experiment B training data the fractional range is 0.42. In contrast, for the filtered Experiment A training data, the fractional range is 0.012. The greater fractional range for Experiment B could be due to more variability in the Experiment B data and fewer features to analyze (compared to the Experiment A). Another possibility is that Experiment B results determined from the unfiltered data are significantly affected by outliers, whereas the Experiment A results are not significantly affected by outliers.

We note that for Experiment A, for $k > 0$, the estimate of the width (σ) of the k -photon energy deposit feature decreases as k increases (see Table 1). However, for Experiment B, σ is larger for the 3-photon energy deposit feature compared to the 2-photon energy photon feature (see Table 3). Perhaps this can be attributed to randomness. Or perhaps there are more outliers in the Experiment B data compared to the Experiment A data. More study is required to resolve this question.

D. Feature set selection

In our feature set selection method (see Section 3.D), the training and test data are the same size. In other applications, one might

consider other splitting schemes. For instance, suppose that the test data were split into J chunks where each chunk corresponds to a non-overlapping time interval. One might consider this splitting scheme for investigation of temporal variations in θ_{dep} . The size of training data might be equated to the size of each of the J chunks of test data.

E. Future research

Unlike the method in [14], our mixture model approach for determination of θ_{dep} does not yield an estimate of the number of photons deposited by any particular pulse. Instead, our mixture model yields an estimate of the expected number of photons in any laser pulse. The task of estimation of the number of deposited photons and the task of estimation of the expected number of deposited photons are different tasks. However, development of method to estimate the number of photons deposited for any particular pulse based solely on the observed pulse height, the estimate of θ_{dep} , and the mixture model parameters is a worthy research topic. One might construct a mixture model based on other summary statistics such as the area of the pulse for a selected time interval or the inner product of a template and the pulse. How well such alternative implementations perform relative to our implementation is a worthy research topic.

In our experience, the optimum filtering approach [22] yields features with shapes that deviate slightly from a Gaussian shape due, in part, to stray photons. Hence, the methods we developed here for quantification of uncertainty for both random and systematic effects and feature selection methods might be applied to analyses based on optimal filtering provided that one identifies a feature shape pdf that would play the role that the Gamma pdf played in our study.

6. SUMMARY

We presented a mixture model method to determine the expected number of photons that deposit energy in a TES per pulse (θ_{dep}) of a laser. The ratio of this estimate and an additional measurement of the expected number of photons created per pulse enables one to calibrate a TES. The choice of features analyzed affects results. From training data, we selected the optimal set of features for analysis according to an uncertainty minimization criterion. Based on the feature set determined from the training data, we determined θ_{dep} and its associated uncertainty from test data that is independent of the training data. Our reported uncertainty for our determination of θ_{dep} accounts for random measurement error, ambiguity in the mathematical form of the mixture model, and possible imperfection in our feature set selection method.

7. ACKNOWLEDGMENTS

Contributions from staff of NIST, an agency of the US Government, are not subject to copyright in the US.

Funding

NIST internal funds.

Disclosures

The authors declare no conflicts of interest.

Data Availability

Data underlying the results presented in this paper are not

publicly available at this time but may be obtained from the authors upon reasonable request.

REFERENCES

1. B. Cabrera, R. Clarke, P. Colling, A. Miller, S. Nam, and R. Romani, "Detection of single infrared, optical, and ultraviolet photons using superconducting transition edge sensors," *Appl. Phys. Lett.* **73**, 735–737 (1998).
2. J. N. Ullom and D. A. Bennett, "Review of superconducting transition-edge sensors for x-ray and gamma-ray spectroscopy," *Supercond. Sci. Technol.* **28**, 084003 (2015).
3. T. Gerrits, A. Migdall, J. C. Bienfang, J. Lehman, S. W. Nam, J. Splett, I. Vayshenker, and J. Wang, "Calibration of free-space and fiber-coupled single-photon detectors," *Metrologia* **57**, 015002 (2019).
4. A. E. Lita, A. J. Miller, and S. W. Nam, "Counting near-infrared single-photons with 95% efficiency," *Opt. Express* **16**, 3032–3040 (2008).
5. A. Avella, G. Brida, I. P. Degiovanni, M. Genovese, M. Gramegna, L. Lolli, E. Monticone, C. Portesi, M. Rajteri, M. L. Rastello *et al.*, "Self consistent, absolute calibration technique for photon number resolving detectors," *Opt. express* **19**, 23249–23257 (2011).
6. G. Brida, L. Ciavarella, I. P. Degiovanni, M. Genovese, L. Lolli, M. G. Mingolla, F. Piacentini, M. Rajteri, E. Taralli, and M. G. Paris, "Quantum characterization of superconducting photon counters," *New J. Phys.* **14**, 085001 (2012).
7. D. Fukuda, G. Fujii, T. Numata, K. Amemiya, A. Yoshizawa, H. Tsuchida, H. Fujino, H. Ishii, T. Itatani, S. Inoue, and T. Zama, "Titanium-based transition-edge photon number resolving detector with 98% detection efficiency with index-matched small-gap fiber coupling," *Opt. Express* **19**, 870–875 (2011).
8. G. McLachlan and D. Peel, *Finite Mixture Models*, Wiley Series in Probability and Statistics (Wiley, 2004).
9. G. J. McLachlan, S. X. Lee, and S. I. Rathnayake, "Finite mixture models," *Annu. review statistics its application* **6**, 355–378 (2019).
10. P. C. Humphreys, B. J. Metcalf, T. Gerrits, T. Hiemstra, A. E. Lita, J. Nunn, S. W. Nam, A. Datta, W. S. Kolthammer, and I. A. Walmsley, "Tomography of photon-number resolving continuous-output detectors," *New J. Phys.* **17**, 103044 (2015).
11. G. Gillett, "A hardware signal processor for transition edge sensors," Ph.D. thesis, The University of Queensland (2017).
12. Z. H. Levine, B. L. Glebov, A. L. Pintar, and A. L. Migdall, "Absolute calibration of a variable attenuator using few-photon pulses," *Opt. express* **23**, 16372–16382 (2015).
13. T. Gerrits, B. Calkins, N. Tomlin, A. E. Lita, A. Migdall, R. Mirin, and S. W. Nam, "Extending single-photon optimized superconducting transition edge sensors beyond the single-photon counting regime," *Opt. Express* **20**, 23798–23810 (2012).
14. Z. H. Levine, T. Gerrits, A. L. Migdall, D. V. Samarov, B. Calkins, A. E. Lita, and S. W. Nam, "Algorithm for finding clusters with a known distribution and its application to photon-number resolution using a superconducting transition-edge sensor," *J. Opt. Soc. Am. B* **29**, 2066–2073 (2012).
15. S. V. Polyakov and A. L. Migdall, "High accuracy verification of a correlated-photon-based method for determining photon-counting detection efficiency," *Opt. Express* **15**, 1390–1407 (2007).
16. A. P. Dempster, N. M. Laird, and D. B. Rubin, "Maximum likelihood from incomplete data via the em algorithm," *J. Royal Stat. Soc. Ser. B (Methodological)* **39**, 1–22 (1977).
17. R Core Team, *R: A Language and Environment for Statistical Computing*, R Foundation for Statistical Computing, Vienna, Austria (2018).
18. T. Benaglia, D. Chauveau, D. R. Hunter, and D. Young, "mixtools: An R package for analyzing finite mixture models," *J. Stat. Softw.* **32**, 1–29 (2009).
19. S. Grinberg, *findPeaks* (GitHub Repository, 2019).
20. B. Efron and R. J. Tibshirani, *An introduction to the bootstrap* (CRC press, 1994).
21. B. N. Taylor and C. E. Kuyatt, *Guidelines for evaluating and expressing*

the uncertainty of NIST measurement results (National Institute of Standards and Technology, NIST TN 1297, 1994).

22. D. Fixsen, S. Moseley, B. Cabrera, and E. Figueroa-Feliciano, "Pulse estimation in nonlinear detectors with nonstationary noise," *Nucl. Instruments Methods Phys. Res. Sect. A: Accel. Spectrometers, Detect. Assoc. Equip.* **520**, 555–558 (2004). Proceedings of the 10th International Workshop on Low Temperature Detectors.